Does k-th Moment Exist?

Hitomi, K.[1] **and Y. Nishiyama**[2]

[1]Kyoto Institute of Technology, Japan
[2]Institute of Economic Research, Kyoto University, Japan
Email: hitomi@kit.ac.jp

Keywords: *Existence of moments, tail index, dependent and heterogeneous data, optimal fraction*

EXTENDED ABSTRACT

Most asymptotic distribution theory used in econometric research relies on moment conditions which carefully control outlier occurrences. It is not unusual in time series analysis to see conditions of the type let all required moments exist. However, in financial and commodity market time series the extent of outlier activity casts doubt on the suitability of such generic moment conditions. Mandelbrot (1963) provided suggestive evidence that even second moments may not exist for this type of data, and he proposed stable distributions with infinite variance as an alternative to finite-variance statistical models. Subsequent research has generally reached the conclusion that second moments of most datasets appear to be finite.

In many empirical works in finance, the values of kurtosis and skewness were statistically tested even though the existence of higher-order, especially fourth or sixth, moments has been studied less extensively.

This paper investigates a statistical testing method for the existence of the k-th moment for dependent, heterogeneous data using the tail index of the distribution function.

Tail index estimation depends for its accuracy on a precise choice of the sample fraction, i.e., the number of extreme order statistics on which the estimation is based.

Our test procedure has two steps. On the first step, we estimate optimal sample fraction that minimizes the mean squared error of Hill's estimator. Then, we test the hypothesis that the k-th moment is exist based on the Hill's estimator.

Results of Monte Carlo simulations show that optimal sample fractions are chosen in average (except for heavily dependent data), size of the test is a slightly higher than a nominal rate and the test has good power for light or moderately dependent data, but the power decreases in heavily dependent case.

1 TAIL INDEX AND HILL'S ESTIMATOR

Suppose $\{X_t\}$ is a sequence of possibly dependent random variable having the same marginal distribution function F, where $\bar{F} = 1 - F$ is regularly varying at ∞, namely there exists an $\alpha > 0$, such that

$$\frac{\bar{F}(tx)}{\bar{F}(x)} \to t^{-\alpha} \quad \text{as } x \to \infty \text{ for all } t > 0 \qquad (1)$$

or equivalently

$$\bar{F}(x) = x^{-\alpha} L(x) \quad x > 0 \qquad (2)$$

for some slowly varying function $L(x)$. α is called tail index of F. Under this, it is easily seen that

$$\bar{F}(b(t)) \sim t^{-1} \quad \text{as } \ t \to \infty$$

where

$$b(t) = F^{-1}(1 - t^{-1}).$$

Briefly speaking, we are assuming that the tail of F could be approximated by a Parate distribution,

$$\bar{F}(x) \sim C x^{-\alpha} \quad \text{as} \quad x \to \infty.$$

An estimator of the tail index could be used for a moment existing test. Since if $k < \alpha$, $E(X^k) < \infty$ and $E(X^k) = \infty$ for $k \geq \alpha$. It is important to be able to accurately estimate the tail index. A range of estimators has been proposed for this task. An intuitive approach to estimation the tail index was conceived by B. Hill (1975). Denoting $z_+ = \max\{z, 0\}$ and $X_{(j)} = X_{(n:j)}$ for the j-th largest value of $X_1, X_2, \ldots, X_n$. Consider a sequence of integers m such that $m \to \infty$ and $m/n \to 0$ as $n \to \infty$. The so called Hill's estimator is defined by

$$\begin{aligned}
\hat{\alpha}_m^{-1} &= \frac{1}{m} \sum_{i=1}^{m} (\ln X_{(i)} - \ln X_{(m+1)}) \\
&= \frac{1}{m} \sum_{t=1}^{n} (\ln X_t - \ln X_{(m+1)})_+ .
\end{aligned}$$

The idea behind Hill's estimator is easily understood. By dominated convergence theorem and integration by parts, the k-th moment of $(\ln X_1 - \ln b(n/m))_+$ is evaluated as (see also Hsing (1991), p1548)

$$\begin{aligned}
& E[(\ln X_1 - \ln b(n/m))_+^k] \\
= & \int_0^\infty P[(\ln X_1 - \ln b(n/m))^k > u] du \\
= & \int_0^\infty \bar{F}(\exp(u^{1/k}) b(n/m)) du \\
= & \bar{F}(b(n/m)) \int_0^\infty \frac{\bar{F}(\exp(u^{1/k}) b(n/m))}{\bar{F}(b(n/m))} du \\
\sim & \frac{m}{n} \int_0^\infty \exp(-\alpha u^{1/k}) du = \frac{m}{n} \frac{k!}{\alpha}.
\end{aligned}$$

For $k = 1$, in particular, $E[(\ln X_1 - \ln b(n/m))_+] = (m/n)\alpha^{-1}$ and $X_{(m+1)}$ estimates $b(n/m)$. Hence the Hill's estimator is essentially a method of moments estimator of α^{-1}.

Hsing (1991), in a seminal paper, proves consistency and established a general distribution limit for $\hat{\alpha}_m^{-1}$ for strong mixing processes. The basic idea is to apply law of large numbers and central limit theorem to $(\ln X_1 - \ln b(n/m))_+$. Hill (2006) extends this results to functionals of near-epoch-dependent on a mixing processes.

2 CHOICE OF SAMPLE FRACTION

Tail index estimation depends for its accuracy on a precise choice of the sample fraction, i.e., the number of extreme order statistics on which the estimation is based.

DuMouchel (1983) suggests that $m \leq 0.1 \times n$ is a good rule. But it does not work well for dependent processes. Table 1 is a result of a small Monte Carlo simulation. The sample size is 1000 and $\{X_t : t = 1, 2, \ldots, 1000\}$ is generated by AR(1) process

$$X_t = 0.9 X_{t-1} + \varepsilon_t,$$

where the ε_t are i.i.d. Student-t with 2 degree of freedom. True values of α^{-1} are 0.5. m is chosen to be 5%, 10%, 15%, 20%, 25% and 30% of the sample size. This table is based on 1000 simulations. The table reports means of Hill's estimator (Hill's est), standard deviations (Std hill's est), and mean squared errors (m.s.e).

From the table, we see the estimator has negative bias for small m ($m = 50, 100$) and positive bias for large m ($m = 200, 250, 300$). The smallest bias and the mean squared error are achieved at $m = 150$. The positive biases for large m are understandable since the order of the bias is $O(m/n)$. Little is known about the bias in small m under serially correlated data. It might be an interesting research topic.

Bootstrap and adaptive selection methods for selecting m in the I.I.D. case are considered in Hall and Welsh (1985), Hall (1990), Drees and Kaufmann (1997), and Danielsson, de Haan, Peng, and de Vries (2001).

Danielsson, de Haan, Peng, and de Vries (2001) suggested the most sophisticated method for estimating mean squared error of the Hill's estimator. They use a sub sampling bootstrap method for estimating bias. Table 2 shows means, standard deviations and mean squared errors of the Hill's estimator for various m. Random variables are drawn from I.I.D. Student-t distribution with 2, 3 and 1 degrees of freedom

Table 1. Hill's estimate under AR(1) a=0.9

A=0.9	n=1000	t d.f. 2				
m	50	100	150	200	250	300
Hill's est	0.4039	0.4623	0.5180	0.5940	0.7094	0.8636
Std hill's est	0.1684	0.1372	0.1199	0.1170	0.1402	0.2272
m.s.e	0.0376	0.0202	0.0147	0.0225	0.0635	0.1838

and Parate distribution with $\alpha = 3/2$ and $5/2$. The smallest mean squared error is marked by asterisk. The smallest mean squared error is achieved at $m = 50$ for the Student-t distribution with 2 and 3 degree of freedom, and at $m = 150$ for the Student-t distribution with 1 degree of freedom (Cauchy distribution). For Parate distribution, optimal m which minimizes the mean squared error is the largest m, since the Hill estimator that uses all data is the maximum likelihood estimator of the Parate distribution.

Table 3 is a result of Danielsson, de Haan, Peng, and de Vries (2001) method. It reports means of selected m, standard deviations of selected m, means of the Hill's estimator and standard deviations of the Hill's estimator. We could see that the Danielsson et al.'s method choose appropriate m.

If we apply their method to dependent data, it choose too small m. Table 4 reports means, standard deviations and mean squared errors of the Hill's estimator for AR(1) data generating process with Student-t errors. AR coefficients are chosen to be 0.6 and 0.9 (corresponding to a=0.6 and a=0.9 in the table). Compared to the I.I.D. case, we could see that the optimal m is increasing. Table 5 reports selected m by Danielsson et al's method and means of the Hill's estimator. Selected m's are too small compared to optimal m's and it generates biases to the estimator.

3 A MEAN SQUARED ERROR ESTIMATOR AND A TESTING PROCEDURE

Our test procedure has two steps. On the first step, we estimate optimal m which minimizes the mean squared error of the Hill's estimator. Then, we test the hypothesis that k-th moment is exist based on the Hill's estimator. It is equivalent to test

$$H_0: \quad \alpha^{-1} < \frac{1}{k}$$
$$H_1: \quad \alpha^{-1} \geq \frac{1}{k}.$$

For minimizing the mean squared error of the Hill's estimator, we need to estimate the variance and the small sample bias of the Hill's estimator.

If the data are I.I.D, Hall (1982) shows the variance of the Hill's estimator is α^{-2}. In general dependent data case, an analytical expression of the variance is not available.

The Hill's estimator is basically a moment estimator of $E[(\log X_1 - \log b(n/m))_+]$, however, we could apply a block bootstrap method for estimating the variance of the Hill's estimator.

For evaluating small sample bias, we should worked on more specific tail probability. Suppose the tail of the distribution function F could be approximated as

$$\bar{F}_t(x) = cx^{-\alpha}(1 + x^{-\alpha}).$$

The following result is Corollary 2 in J. B. Hill (2007).

Assumption A $\{X_t\}$ satisfies (1) for some $\alpha > 0$. For some positive measurable $g : R_+ \to R_+$,

$$L(\lambda x)/L(x) - 1 = O(g(x)) \quad \text{as } x \to \infty.$$

The function g has bounded increase: there exists $0 < D$, $z_0 < \infty$ and $\tau < 0$ such that $g(\lambda z)/g(z) \leq D\lambda^{\tau}$ some for $\lambda \geq 1$ and $z \geq z_0$. Specifically, $\{m_n\}_{n\geq 1}$, $\{b_{m_n}\}_{n\geq 1}$, and $g(.)$ satisfy

$$\sqrt{m_n} g(b_{m_n}) \to 0, \quad \text{where } m_n \to \infty, m_n = o(n).$$

Assumption B $\{X_t\}$ is L_2-E-NED with size 1/2 on E-Mixing Process $\{\epsilon_t\}$. The base $\{\epsilon_t\}$ is E-Uniform Mixing with size $r/(2(r-1))$ for some $r \geq 2$, or E-Strong Mixing with size $r/(r-2)$ for some $r > 0$.
Remark: See J. B. Hill (2007) for definitions of L_2-E-NED, E-Mixing and E-Uniform Mixing.

Lemma Suppose

$$\bar{F}_t(x) = cx^{-\alpha}(1 + x^{-\alpha}), \quad \alpha, c > 0,$$

and $m_n = o(n^{2/3})$. Under Assumptions A and B

$$\begin{aligned} B_{m_n} &\equiv E[\sqrt{m_n}(\hat{\alpha}_{m_n}^{-1} - \alpha^{-1})] \\ &= \sqrt{m_n}\alpha^{-1} \times \frac{1}{2} \times b(n/m)^{-\alpha} + o(1) \\ &= o(1). \end{aligned}$$

Note: For small samples and any process satisfying (2), this estimator of the bias is at best a rough approximation of the true bias. Pareto random variables, for example, this estimator over-estimates the bias.

$b(n/m)$ is easily estimated by $X_{(m+1)}$. Our estimator of the bias is

$$\hat{B}_m = \sqrt{m}\hat{\alpha}_m^{-1} \times \frac{1}{2} \times X_{(m+1)}^{-\hat{\alpha}_m}$$

and the estimator of the optimal m is

$$\hat{m} = \min_m \{\hat{\sigma}_m^2 + \hat{B}_m\},$$

where $\hat{\sigma}_m^2$ is a block bootstrap estimator of the variance of $\hat{\alpha}_m^{-1}$. Our test statistics is usual t-statistics that uses the Hill's estimator evaluated at $\hat{m}$,

$$t = \frac{\hat{\alpha}_{\hat{m}}^{-1} - 1/k}{\sqrt{\hat{\sigma}_{\hat{m}}^2}}.$$

Asymptotic normality of the test statistics has not yet been proved. We only present a small Monte Carlo simulation result at this time. The samples are generated as follow. We draw random samples of I.I.D. innovations $\{\epsilon_t\}$ from Student-t distribution 2,3 or 1 degree of freedom. It implies true $\alpha^{-1} = 0.5, 0.33$ or 1.0 correspondingly. The number of sample size is fixed to 1000. We simulate AR(1) processes using $\{\epsilon_t\}$,

$$X_t = a \times X_{t-1} + \epsilon_t, \quad a \in \{0.3, 0.6, 0.9\}.$$

500 processes are generated for each cases. Our testing hypothesis is that the variance is finite or not,

$$\begin{aligned} H_0 : &\quad \alpha^{-1} < 1/2 \\ H_1 : &\quad \alpha^{-1} \geq 1/2. \end{aligned}$$

Results are summarized in Table 6. Table 6 reports means of $\hat{m}$, s.t.d of $\hat{m}$, means of $\hat{\alpha}_{\hat{m}}^{-1}$, s.t.d of $\hat{\alpha}_{\hat{m}}^{-1}$ and rejection rates of H_0 based on 5% one side nominal critical value for each AR coefficient and degree of freedom of t-distribution.

Compared to Table 4, we could see that appropriate m's are chosen in average. It implies the biases are not so large at least $a = 0.3$ and 0.6.

Rejection rate of H_0 on Student-t with 2 degree of freedom shows size of the test. It shows a tendency to over rejection (about 2-3%). The power of the test is not so bad. The rejection rate of Student-t with 1 degree of freedom is almost one for $a = 0.3$ and 0.6. But it decreases to 60% when $a = 0.9$. Probably, this power loss is caused by relatively large bias at $a = 0.9$.

4 CONCLUSION

This paper proposes a statistical testing method for the existence of the k-th moment for dependent, heterogeneous data using the tail index of the distribution function.

Our test procedure has two steps. On the first step, we estimate optimal m which minimizes the mean squared error of the Hill's estimator. Then, we test the hypothesis that the k-th moment is exist based on the Hill's estimator.

The results of Monte Carlo simulations show that optimal m's are chosen in average (except for heavily dependent data), the size of the test is a slightly higher than a nominal rate and the test has good power for light or moderately dependent data but the power decreases in heavily dependent case.

5 REFERENCES

DANIELSSON, J., L. DE HAAN, L. PENG, AND C. G. DE VRIES (2001): "Using a Bootstrap Method to Choose the Sample Fraction in Tail Index Estimation," *Journal of Multivariate Analysis*, 76(2), 226–248.

DE HAAN, L., AND S. I. RESNICK (1980): "A Simple Asymptotic Estimate for the Index of a Stable Distribution," *Journal of the Royal Statistical Society. Series B*, 42(1), 83–87.

DEKKERS, A. L. M., AND L. D. HAAN (1989): "On the Estimation of the Extreme-Value Index and Large Quantile Estimation," *The Annals of Statistics*, 17(4), 1795–1832.

DREES, H., AND E. KAUFMANN (1998): "Selecting the optimal sample fraction in univariate extreme value estimation," *Stochastic Processes and their Applications*, 75(2), 149–172.

DUMOUCHEL, W. H. (1983): "Estimating the Stable Index α in Order to Measure Tail Thickness: A Critique," *The Annals of Statistics*, 11(4), 1019–1031.

HALL, P. (1982): "On Some Simple Estimates of an Exponent of Regular Variation," *Journal of the Royal Statistical Society. Series B (Methodological)*, 44(1), 37–42.

——— (1990): "Using the Bootstrap to Estimate Mean Squared Error and Select Smoothing Parameter in Nonparametric Problems," *Journal of Multivariate Analysis*, 32(1), 177–203.

HALL, P., AND A. H. WELSH (1985): "Adaptive Estimates of Parameters of Regular Variation," *The Annals of Statistics*, 13(1), 331–341.

HILL, B. M. (1975): "A Simple General Approach to Inference About the Tail of a Distribution," *The Annals of Statistics*, 3(5), 1163–1174.

Table 2. Mean and M.S.E. of Hill's Estimator (I.I.D.)

	m	50	100	150	200	250	300	400
t dist d.f. 2	mean	0.5431	0.5883	0.6463	0.7216	0.8103	0.9340	1.3930
$\alpha^{-1} = 0.5$	s.t.d	0.0757	0.0539	0.0479	0.0459	0.0493	0.0607	0.1267
	m.s.e	0.0076*	0.0107	0.0237	0.0512	0.0987	0.1920	0.8136
t dist d.f. 3	mean	0.4081	0.4693	0.5377	0.6189	0.7188	0.8433	1.3092
$\alpha^{-1} = 1/3$	s.t.d	0.0532	0.0419	0.0408	0.0408	0.0451	0.0543	0.1188
	m.s.e	0.0084*	0.0202	0.0434	0.0832	0.1506	0.2630	0.9665
t dist d.f. 1	mean	1.0103	1.0252	1.0578	1.1025	1.1701	1.2676	1.6680
$\alpha^{-1} = 1$	s.t.d	0.1475	0.0994	0.0824	0.0754	0.0694	0.0726	0.1182
	m.s.e	0.0219	0.0105	0.0101*	0.0162	0.0337	0.0769	0.4601
	m	50	100	200	400	500	750	
Parate	mean	0.6672	0.6678	0.6651	0.6657	0.6664	0.6658	
$\alpha^{-1} = 0.66$	s.t.d	0.0947	0.0662	0.0454	0.0320	0.0298	0.0241	
	m.s.e	0.0090	0.0044	0.0021	0.0010	0.0009	0.0006*	
Parate	mean	0.3998	0.4002	0.3987	0.3994	0.4005	0.3999	
$\alpha^{-1} = 0.40$	s.t.d	0.0567	0.0384	0.0280	0.0203	0.0184	0.0147	
	m.s.e	0.0032	0.0015	0.0008	0.0004	0.0003	0.0002*	

Table 3. Danielson et al. (2001) Method

	Student-t			Parate	
	d.f. 2	d.f. 3	d.f. 1	$\alpha^{-1} = 0.66$	$\alpha^{-1} = 0.4$
m	54.6800	33.69	119.3720	864.6080	874.8760
s.t.d of m	43.1062	28.3625	92.72	183.7278	184.2408
mean of Hill's est	0.5187	0.3673	1.0579	0.6623	0.3966
s.t.d of Hill's est	0.1109	0.0870	0.8353	0.0406	0.0191

Table 4. Mean and M.S.E. of Hill's Estimator (AR(1))

		M	50	100	150	200	250	300	400
t d.f. 2	a=0.6	mean	0.4816	0.5285	0.5800	0.6569	0.7516	0.8764	1.3661
		s.t.d	0.1006	0.0699	0.06	0.0602	0.06	0.0859	0.2388
		m.s.e	0.0105	0.0057*	0.0105	0.0283	0.0675	0.1491	0.8071
	a=0.9	mean	0.4039	0.4623	0.5180	0.5940	0.7094	0.8636	1.3237
		s.t.d	0.1684	0.1372	0.1199	0.1170	0.1402	0.2272	0.6311
		m.s.e	0.0376	0.0202	0.0147*	0.0225	0.0635	0.1838	1.0769
t d.f. 3	a=0.6	mean	0.3514	0.4104	0.4846	0.5609	0.6690	0.8034	1.2892
		s.t.d	0.0597	0.0453	0.0467	0.0483	0.0602	0.0803	0.2346
		m.s.e	0.0039*	0.0080	0.0250	0.0541	0.1163	0.2274	0.9686
	a=0.9	mean	0.2838	0.3494	0.4263	0.5179	0.6370	0.7740	1.2714
		s.t.d	0.0902	0.0745	0.0749	0.0865	0.1175	0.1740	0.6110
		m.s.e	0.0106	0.0058*	0.0142	0.0415	0.1060	0.2245	1.2533
t d.f. 1	a=0.6	mean	0.9655	1.0030	1.0354	1.0894	1.1697	1.2765	1.7133
		s.t.d	0.2581	0.1955	0.1688	0.1423	0.1271	0.1335	0.2658
		m.s.e	0.0678	0.0382	0.0297	0.0282*	0.0449	0.0943	0.5794
	a=0.9	mean	0.8051	0.9286	0.9890	1.0771	1.1810	1.2952	1.6095
		s.t.d	0.3671	0.3413	0.2985	0.3092	0.2806	0.3171	0.7698
		m.s.e	0.1728	0.1216	0.0892*	0.1015	0.1115	0.1877	0.9641

Table 5. Danielsson et al.'s method (AR(1))

		t d.f. 2	t d.f. 3	t d.f. 1
a=0.6	m	50.06	31.08	105.80
	mean of Hill's	0.44	0.31	1.00
a=0.9	m	46.79	30.71	78.37
	mean of Hill's	0.35	0.22	0.73

Table 6. Finiteness of variance test

	AR coef	a=0.3	a=0.6	a=0.9
t dist d.f. 2	mean of $\hat{m}$	68.30	93.80	115.90
	s.t.d of $\hat{m}$	24.11	31.36	58.56
	mean of $\hat{\alpha}_{\hat{m}}^{-1}$	0.5162	0.5091	0.4579
	s.t.d of $\hat{\alpha}_{\hat{m}}^{-1}$	0.0699	0.0827	0.1625
	rejection rate of H_0	0.0820	0.0700	0.0620
t dist d.f. 3	mean of $\hat{m}$	59.50	79.20	87.20
	s.t.d of $\hat{m}$	19.63	79.20	45.06
	mean of $\hat{\alpha}_{\hat{m}}^{-1}$	0.3913	0.3747	0.3197
	s.t.d of $\hat{\alpha}_{\hat{m}}^{-1}$	0.0519	0.3747	0.1077
	rejection rate of H_0	0.0000	0.0000	0.0080
t dist d.f. 1	mean of $\hat{m}$	84.90	110.20	136.80
	s.t.d of $\hat{m}$	32.85	53.40	88.33
	mean of $\hat{\alpha}_{\hat{m}}^{-1}$	0.9767	0.9627	0.8755
	s.t.d of $\hat{\alpha}_{\hat{m}}^{-1}$	0.1590	0.1943	0.3386
	rejection rate of H_0	0.9900	0.9800	0.5960

HILL, J. B. (2006): "On Tail Index Estimation for Dependent, Heterogenous Data," mimeo.

——— (2007): "Kernel Methods for Small Sample and Asymptotic Tail Inference of Dependent, Heterogeneous Data," mimeo.

HSING, T. (1991): "On Tail Index Estimation Using Dependent Data," *The Annals of Statistics*, 19(3), 1547–1569.

KEARNS, P., AND A. PAGAN (1997): "Estimating the density Tail Index for Financial Time Series," *Review of Economics and Statistics*, 79(2), 171–175.

LORETAN, M., AND P. C. PHILLIPS (1994): "Testing the covariance stationarity of heavy-tailed time series," *Journal of Empirical Finance*, 1(2), 211–248.

MANDELBROT, B. (1963): "The Variation of Certain Speculative Prices," *The Journal of Business*, 36(4), 394–419.

PICKANDS, J. (1975): "Statistical Inference Using Extreme Order Statistics," *The Annals of Statistics*, 3(1), 119–131.